

今天主要是继续写综述，先从第二章开始，个人觉得这章是目前研究比较薄弱的环节，未来可能会有较大发展，特别是最近出现了很多更先进的交互设备，那么必然交互界面也会发生变化。第二章介绍查询可视化，也就是如何进行可视化输入，让系统更准确地理解用户意图同时简化用户的操作。今天写了基于分类的查询可视化的两小节，分别是层次分类和层面分类，都是比较经典的输入方式，但是相关的研究非常多，有的是研究可视化方法，有的是对这些方法的评测。例如，**SuperBook** 按书籍的目录结构组织搜索结果，评测表明这种用户界面不仅能加快简单的信息检索任务而且能提高搜索的精度。**Allen** 调研了类似的两个数字图书馆界面，**Dewey Decimal System** 以及 **ACM Computer Reviews** 系统，表示使用层次分类能有效地组织文档集。图书搜索系统 **Cha-Cha** 根据网站的结构自动将搜索结果组织成层次结构，便于用户查找难以发现的信息。类似地 **WebTOC** 将网站结构组织成树型结构，可以展开或回缩，同时嵌套直方图显示每个文档文本、图片、音频的比例。**Flamenco** 使用多级层面搜索界面对固定大小的文档集进行搜索，用户选择任何层面都会对搜索结果产生明显改变。有文献针对 **Flamenco** 接口和关键词搜索接口比较两者的实用性，分别对艺术和建筑图执行结构化开放式的探索任务。使用 **Flamenco** 用户发现相关图片的成功率更高而且体现出更强的主观性。这两类可视化方法是目前应用最成功的，尤其是层面分类目前广泛应用于购物网站的搜索系统中。

在写这部分内容时想到，是否可以先抛开数据分析专门开展交互界面的研究，例如针对层次多分类数据定义一套图形规则，根据规则将用户输入的图形转换成 **Data Cubes** 上的操作，进而实现数据查询。在实际应用中确实存在很多难以操作的查询，例如“3 月份家电类的哪种商品在哪个网站销售量最高”，用户即使使用最先进的可视分析软件 **tableau**，也要多次交互。首先要人工找到家电分类，然后逐一比较哪种商品销售量最高，再根据最高的销售量找到对应的网站。如果再增加一个条件，如“3 月份家电类的哪种商品在哪个网站销售量最高并且销售金额最大”。同时比较多种属性，对用户来说是非常痛苦的，因为在搜索时人的注意力通常只能关注于少量的特征。如果在输入时精确传达用户意图，那么最后需要留给用户选择的只有少量不确定的特征了。上例中，“3 月、家电类”是可以确定的，那么最后的结果中，就没有必要体现“3 月和家电类”，只要表示商品、网站、销售量即可，减少了维度，在可视化界面中减轻了用户认知负担。

另外，今天阅读了 **Tamara Munzner** 的新书其中关于可视化编码原则的部分，写得非常清晰、易懂。不过个人觉得最精彩的还是前面关于抽象的阐述，把可视化当作任务，指导如何开展可视化过程，非常实用。遗憾的是中间缺少了部门章节，没有说明具体的任务抽象和数据抽象。